



Cite this: *Mol. BioSyst.*, 2016,
12, 2965

Received 18th May 2016,
Accepted 18th July 2016

DOI: 10.1039/c6mb00388e

www.rsc.org/molecularbiosystems

Cellular identity at the single-cell level

Ahmet F. Coskun,^{*a} Umut Eser^b and Saiful Islam^c

A single cell creates surprising heterogeneity in a multicellular organism. While every organismal cell shares almost an identical genome, molecular interactions in cells alter the use of DNA sequences to modulate the gene of interest for specialization of cellular functions. Each cell gains a unique identity through molecular coding across the DNA, RNA, and protein conversions. On the other hand, loss of cellular identity leads to critical diseases such as cancer. Most cell identity dissection studies are based on bulk molecular assays that mask differences in individual cells. To probe cell-to-cell variability in a population, we discuss single cell approaches to decode the genetic, epigenetic, transcriptional, and translational mechanisms for cell identity formation. In combination with molecular instructions, the physical principles behind cell identity determination are examined. Deciphering and reprogramming cellular types impact biology and medicine.

Introduction

Multicellular organismal life starts from a single cell and experiences significant physiological and molecular changes during development under dynamic environmental stimuli.

A human body is formed based on a single zygote cell that creates more than 10^{13} specialized cells with unique subcellular structures and complex functions after extensive cellular division.¹ Although individual cells in an organism share almost indistinguishable genomic material, cellular proliferation and differentiation lead to remarkable diversity that makes up different parts of organisms such as blood formation, neural development, and organ growth (Fig. 1). Regulatory mechanisms are considered to facilitate cellular specialization in development across the central dogma covering deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and protein conversions.² Genomic DNA content, messenger RNA expression, and protein

^a Division of Chemistry and Chemical Engineering, California Institute of Technology, California, USA. E-mail: ahmet@caltech.edu

^b Department of Genetics, Harvard Medical School, Massachusetts, USA. E-mail: eser@genetics.med.harvard.edu

^c Department of Genetics, Stanford Genome Technology Center, Stanford University, Stanford, USA. E-mail: mdsaiful@stanford.edu



Ahmet F. Coskun

degree from the University of California, Los Angeles, and a bachelor degree from Koc University, Turkey.

Ahmet F. Coskun is a Research Fellow at the Division of Chemistry and Chemical Engineering at the California Institute of Technology. He is a recipient of the Burroughs Wellcome Fund Career Award at the Scientific Interface. He researches systems biology, precision medicine, and biophotonics. His current research centers on single cell analysis in biology and medicine using interdisciplinary quantitative tools. He holds a PhD



Umut Eser

Department of Genetics at the Harvard Medical School. His research interests include evolution of transcription regulation, deep learning applications on genomics and epigenomics, and statistical representation of complex data.

Umut Eser was awarded a gold medal in the 33rd International Physics Olympiad. He earned his BS degree in Physics from the Middle East Technical University in 2006 and received a PhD in Applied Physics at Stanford University in 2013. His doctorate thesis focuses on understanding the design principles of cellular decision making, particularly cell cycle commitment. Currently, he continues his postdoctoral studies in the

abundance have been weakly correlated within biological processes in cells.^{3,4} Besides, the molecular patterns that a cell expresses are very dynamic and subject to change over time. Thus, genetic information transfer from DNA to protein exhibits significant heterogeneity and stochasticity, resulting in a variety of cell fates and phenotypes that define specific cell types.

Genome-wide DNA, RNA, and protein profiling methods have elucidated potential regulatory mechanisms for cellular differences.^{5,6} However, these approaches have been limited to the ensemble average data that combine information from a group of cells. The heterogeneity in cellular populations is resolved by analyzing the regulatory elements in individual cells. Thus, single cell technologies have received considerable attention to explore population architecture in organisms consisting of various cellular subtypes. For example, emerging single-cell RNA sequencing (RNA-seq) and single-cell quantitative polymerase chain reaction (q-PCR) techniques have mapped out the molecular states of individual cells in both healthy and diseased model systems. While cells have been classified based on gene expression data in certain organisms, understanding global regulation that gives rise to cellular identities requires sophisticated analysis of epigenome, genome, and proteome landscapes together in spatial and temporal domains.

Here, we discuss molecular coding mechanisms for cellular identity formation especially covering genomic, epigenomic, transcriptional, and translational regulation in cellular development. We provide single cell approaches that have shed light on some of these codes in development. In particular, we review the observed cellular states using available single cell profiling techniques. We then highlight the need for further advances in single cell techniques. In addition to molecular programming, we discuss the physical mechanisms leading to cell type formation. Finally, we overview some of the applications of cell identity studies in biology and medicine.



Saiful Islam

Saiful Islam has contributed significantly during his PhD at Karolinska Institutet to the development of single cell transcriptomics methods including Single-cell Tagged Reverse Transcription (STRT) for single cell transcriptomics and a Unique Molecular Identifier (UMI) based molecule-counting method at the single-cell level. He is currently working as a postdoctoral research scholar in the Genetics Department at Stanford University and applies single cell analysis to further understand the biology of clonal T-cell expansion in response to pathogens. In addition, he is working on developing a new single cell method to deal with very small cells containing cell walls.

and applies single cell analysis to further understand the biology of clonal T-cell expansion in response to pathogens. In addition, he is working on developing a new single cell method to deal with very small cells containing cell walls.

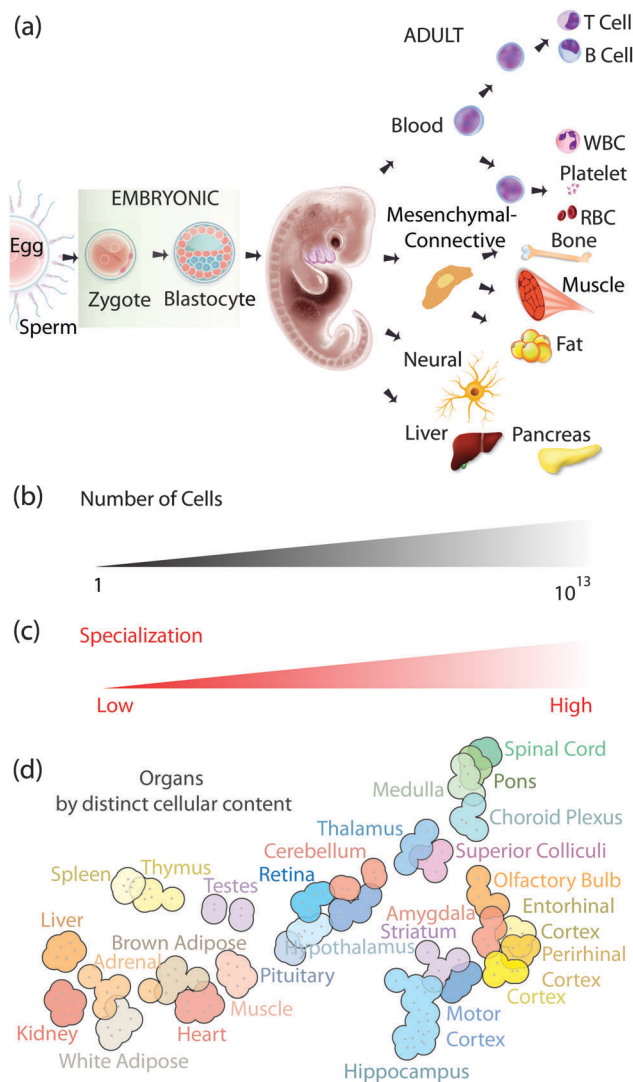


Fig. 1 (a) Single zygote cell produces cellular diversity in a multicellular organism in development. (b) One cell creates up to trillions of functional cells. (c) A series of differentiation events create highly specialized cells. (d) Microarray data show different cell types from gene expression analysis of 80 tissue samples. Colours represent distinct tissue types and each grey dot denotes a gene expression data point for each tissue sample.

Molecular coding

Recent advances in whole genome sequencing have revealed that humans share 18–95% of a common genome with other organisms such as worms, fruit flies, zebrafish, plants, dogs, and mice.^{7–10} Similarities in their DNA sequences help these organisms perform conserved cellular functions such as growth and movement, among others. However, the DNA sequence differences are primarily responsible for making each organism unique with a special molecular make up and physical structure. Any two human beings are identical in 99.9% of their DNA content.¹¹ Despite this high degree of similarity, humans exhibit significant differences between individuals in their physical appearances and capabilities. Person-to-person variations may arise from small sections of our DNA that contain only 0.1% of

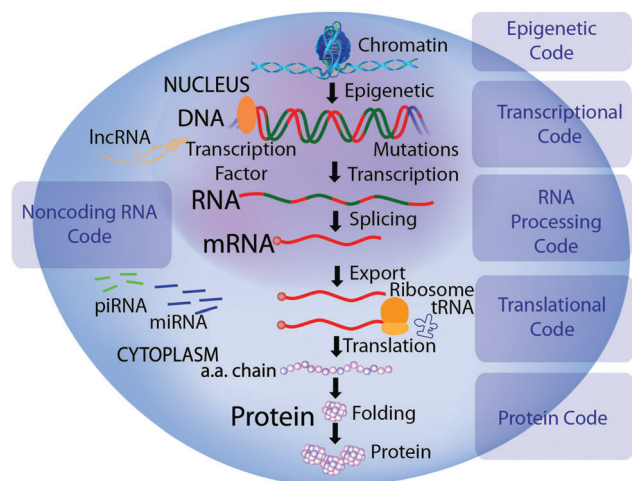


Fig. 2 Cells are coded at different levels of the central dogma to create cellular diversity.

the entire 6 billion base pairs corresponding to a significant amount of “6 million” base pairs per cell.

In addition, mutations in DNA contribute to this diversity. In humans, DNA polymerase fidelity is estimated to be about 10^{-8} per genomic site per generation. For a 6.6×10^9 base-pair diploid genome, on average 60 mutations occur per cell division.^{12–14} Everyone naturally contains hundreds of mutations and most of them persist during the life span. Such a significant mutation rate has an impact on human conditions especially for diseases. In the human body, the DNA is almost identical in every cell, but still many specialized cell types have been produced within different tissues and organs (Fig. 1). To shed light on cell identity formation, the following section will discuss molecular coding mechanisms (Fig. 2) across the central dogma of biology that lead to cellular specialization.

Genetic code

Cellular DNA contains genetic instructions for human traits and transmits to the next generations through inheritance. All of the cells inherit the same DNA from the germline cells, and thus, the cells within an organism have almost the same DNA sequences (exceptions are discussed in the next section). These cells make use of an instruction manual, known as the genetic code, to read the nucleotide sequence of a gene and construct a corresponding protein structure. In particular, the code comprises a direct mapping of every triplet of nucleotides in a nucleic acid sequence to a single amino acid.¹⁵ While the majority of genes are encoded with exactly the canonical or standard genetic code, there are many other variant or non-canonical codes. Therefore, the genetic code is not universal and has continued to expand its coding capacity throughout evolution.^{16,17} Such naturally evolved genetic code in cells may play an important role in the proteome balance, which might be critical for survival functions. Despite recent evidence in genetic code modifications, the cell identity formation is tightly linked to the activation of subsets of genes through a presumed universal genetic code. Simply, the DNA sequences are conserved,

but selective and specific use of DNA gives rise to different cell types in development.

Genomic code

Different cells in an organism have minor differences in DNA sequences due to genomic aberrations including somatic mutations, single-nucleotide polymorphisms (SNPs), deletions, structural variants, and copy number variations (CNVs), and genetic mosaicism.¹⁸ Additional genomic differences are prevalent in immune cell types such as B-cells that modify a small part of DNA to produce antibodies and some organisms with programmed DNA elimination.¹⁹ While normal development does not phenotypically change with these small genomic variations, cancer formation is partially attributed to these mutations.²⁰ Whole genome amplification and sequencing methods have been developed to access these genomic variations in diseases such as neurological disorders.²¹

Chromosomal conformation code

The human genome is hierarchically organized in three dimensions (3D) leading to a fractal globular shape in the nucleus.²² Different juxtapositions of transcription factories, enhancers, and promoters interact with each other in physical space, giving rise to differential expression patterns.^{23,24} Such variation in chromosomal conformation is not only found among different organisms, but is also observed among different tissues of the same organism. Distinct cell types exhibit unique 3D chromatin structures and genome topology. Specifically, topologically associated domains (TADs) control physical interactions of genomic elements (enhancers and promoters) to regulate genomic activity. TADs determine distinct 3D chromosome structures based on the cell type origin.²⁵ Altering the structure of TADs and boundaries can lead to diseases.²⁴ Notably, 3D genome structure modifications play a central role in cell identity determination for other mechanisms such as transcriptional and epigenetic control, the details of which will be expanded in the following sections.

Transcriptional code

In the quest for determining their fates, cells combine cascaded messages coming from signaling networks and transcriptional regulators to decide which set of genes to express. Specifically, cell identity is formed through the interplay between transcriptional factors and environmental stimulation such as *Notch* signaling in immune cell development or *Shh* signaling in central nervous system development.^{26–29} Transcription factors play a role as activators or repressors of gene expression in developing organs. Cell types express a specific set of transcription factors as a barcode identifier for a cell. Thus, combinatorial transcription regulation lies at the heart of developmental processes in organisms. For instance, an appropriate combination of transcription factors (Brn2, Ascl1, Myt1l, and Neurod1) is sufficient to reprogram fibroblasts to other cell types such as neurons.³⁰

In the DNA–protein association data by Chromatin immunoprecipitation followed by sequencing (ChIP Seq), transcription

factors typically bind to *cis*-acting or *trans*-acting regulatory elements known as enhancers to regulate gene expression.^{31,32} Enhancers can be shared by multiple transcriptional factors and exhibit cell type specific distribution. Therefore, transcriptional enhancers contribute to the selection of cell fates. Master transcriptional factors form clusters of enhancers, which are referred to as super-enhancers.^{33,34} Chromosome 3D local structures control super-enhancer driven cell identity formation. Specifically, local chromosome loops contribute to the insulation of neighbouring genes. In addition, cancer cells form super enhancers for oncogenes.³⁵ While the concept of super enhancers is relatively new and requires substantial validation, their emerging role may be important in cell identity both in health and disease.³⁶

Epigenetic code

A complementary mechanism to the interplay of local regulatory DNA sequence and transcription factors, by orchestrating the gene expression through the changing state of the chromatin, is called epigenetics.³⁷ Epigenetic code primarily researches the roles of DNA packaging histone proteins and chemical modification of the nucleotides such as DNA methylation.³⁸ Compared to universal genetic code, epigenetic codes exhibit cell type specific regulation, providing a critical component of cell identity studies.

Histone code is an epigenetic factor that partly changes transcription by chemical modifications to histone proteins.^{39–42} Combinations of covalent histone modifications such as methylation, acetylation, phosphorylation, and ubiquitination recruit other specific proteins to modify the chromatin structure, causing activation or repression of a gene. Enhancers in human genome are significantly marked by histone modifications in distinct cells to enable cell type specific gene expression.⁴³ To maintain the cell identity against variations of the stochastic environment, histone modifications such as H3K4me3 consistently mark the key cell identity genes in specific tissues.⁴⁴ Polycomb-group (PcG) protein complexes catalyse histone modifications to silence transcription in embryonic development.^{45–48} Epigenetic PcG repression plays an important role for stem cells in health and disease such as cancer.

DNA methylation adds to the modularity of epigenetics to affect cell type specific gene expression.⁴⁹ Covalent addition of a methyl group to the cytosine across the genome causes differential regulation in the adjacent genes.⁵⁰ For instance, pancreatic β cell identity is controlled by the repression of an *aristaless*-related homeobox (*Arx*) gene through the DNA methylation mechanism.⁵¹

Non-coding DNA sequences such as transposable elements have an impact on cell type specific transcriptional activity. These mobile DNA pieces move around the DNA genome. In particular, L1 retrotransposition, a mechanism in which reverse transcribed elements are inserted into the genome, exhibits differential regulation in germ cells, stem cells, and neuronal cells.^{52–57}

Another epigenetic modulator is non-coding RNAs in cells. Long non-coding RNAs (lncRNAs) play an important role in reprogramming mammalian cells.^{58–60} Human tissues exhibited

cell specific expression of lncRNAs in β cells. A mesoderm specific *Fendrr* lncRNA provided tissue specific expression in heart and body wall development. lncRNAs appear to be crucial gene expression regulators by affecting nuclear organization and sequestering microRNAs (miRNAs), behaving as miRNA sponges.^{61–65} In addition, microRNAs are short non-coding RNAs that can manipulate the expression of genes to modulate the differentiation status of cell identity.⁶⁶ miRNAs cooperate to silence target mRNAs.⁶⁷ Another class of small RNAs includes Piwi interacting RNAs (piRNAs) that repress the activity of retrotransposons especially in somatic and germline stem cells.^{68,69} piRNAs control lineage determination in multiple cell types within the *Drosophila* ovary.⁷⁰

While the term “code” for epigenetic modifications has been widely used in cell biology, it is still at the level of hypothesis and begs the question for concrete evidence. Current observables have been limited to correlations. For example, H3K36me3 is associated with gene expression at the 3′ end of the genes. However, there are cases where the gene is not expressed and still H3K36me3 was pronounced, or alternatively, the gene is active but H3K36me3 was not detected in cellular machineries. Besides, the majority of these histone marks correlate highly with each other, raising the issue of whether those are determinants; or, just marks of the DNA state; or which ones are the determinants.

RNA code

RNA processing of the nucleotide sequence affects cell identity. Both co-transcriptional and post-transcriptional regulation mechanisms of gene expression are controlled by alternative splicing events. Splicing of a transcript exhibits cell type specific regulation differences to perform specialized functions in differentiated cells.^{71–74} Moreover, an alternatively spliced transcript can produce diverse proteins, which interact with a different set of proteins as if they originated from different genes.⁷⁵ Cells acquire splicing code based on the developmental context (sex, age, function, and organism) to obtain the necessary diverse protein landscape for cellular functions.

RNA editing is another post-transcriptional modification of RNA molecules to change the nucleotides on the transcripts. It is a rare alteration that has been observed in neuronal cell identity development.⁷⁶ In particular, adenosine-to-inosine editing in RNAs creates protein isoforms for creating neural excitability differences.⁷⁷ 14 human cell lines exhibited differences in their RNA editing patterns.⁷⁸ Tissue specific RNA editing was mapped out by a computational tool (GIREMI).⁷⁹ The APOBEC3A gene facilitates RNA editing events in macrophages and monocytes.⁸⁰

Translational code

The ribosomal machinery has classically been considered to act constitutively in the translation of mRNAs to proteins. However, there is increasing evidence to support a specialized ribosome theory that exhibits transcript specific regulation in development.⁸¹ Thus, a ribosome-mediated regulation mechanism⁸² leads to cell identity formation. The ribosomal code suggests

that unique ribosomes with structural and functional differences may prefer translation of subsets of mRNAs in different cell types. Ribosomal compositional variations (composed of 4 ribosomal RNAs and 80 proteins) have recently been measured by mass spectroscopy to reveal differential stoichiometry among ribosomal proteins.⁸³

Protein code

While transcriptional and translational mechanisms encode proteins, the dynamic regulation of proteins to perform specific cellular functions is guided by post-translational modifications such as phosphorylation, among more than 200 others. For instance, cells utilize such modifications of proteins to maintain the body's circadian clock and basic cell division tasks, as well as to avoid cancer formation.^{84–86} Thus, protein modifications are critical for regulatory specificity in differentiated and specialized cells.

Proteins must be folded properly and pass the quality control in a cell.⁸⁷ Aggregation of misfolded proteins causes a loss of cell identity, increasing the prevalence of critical diseases such as cancer. Next, protein–protein interactions experience cell type specificity to obtain distinct cellular identities. For instance, LIM factor Lhx3 binds to LIM cofactor NLI to stimulate interneuron specification in development.⁸⁸ Protein complexes indirectly repress the activity of a transcriptional factor to guide the selection of a neuronal fate.

Single cell decoding

Powerful genome wide DNA/RNA sequencing technologies have successfully mapped out molecular codes for cell identity formation.^{5,6} However, the presented approaches can only provide ensemble data. Sub-population biological studies have been lacking due to these sequencing limitations. The non-uniform distribution of molecular components across the population might be critical to understand how cells reach a fate in a multicellular system (Fig. 3). Besides, gene expression levels exhibit significant variability in cells due to multiple regulatory mechanisms as previously discussed in the molecular coding section.⁸⁹

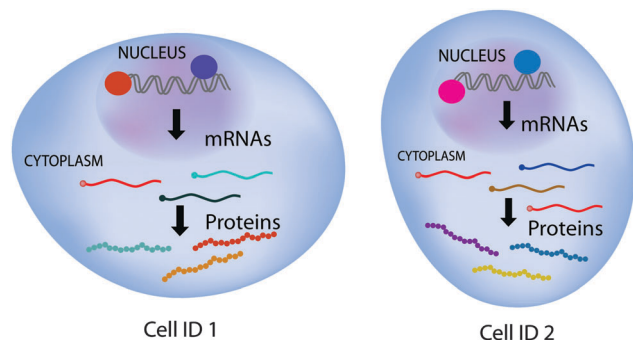


Fig. 3 Distinct cell types have unique physical and molecular signatures. Single cell molecular profiling provides the identity of each individual cell. Despite common DNA sequences, protein (colour filled circles) and genome (grey) interactions create a heterogeneous distribution of RNAs and proteins, creating different cell identities: Cell ID 1 and 2.

Here, we discuss the identification of cellular identity based on single cell molecular profiling techniques from the genome to the proteome scale.

Single cell genomics

Nucleotide variants in the genome lead to different cellular functions. Mainly, somatic mutations cause nucleotide variants that accumulate in the genome of a single-cell over several divisions. Even though somatic variations are routinely prevalent in diseased and normal tissues, the rate of somatic variations is unexpectedly high.^{90,91} These mutations give each cell a unique identity with poorly defined functions. Thus, to decipher the role of dynamic mutations in cellular functions, single cell whole genome sequencing technologies reveal previously obscured variability and complexity in the genetic factors leading to those mutations. Multiple types of variations occur in gene function due to the burst size and frequency in gene expression, as well as cell cycle differences. To access the genome differences of each cell, DNA sequencing of single cell materials requires whole genome amplification methods such as the PCR based method, the multiple annealing and looping based amplification (MALBAC) isothermal method, and multiple displacement amplification (MDA). Microarrays and next generation sequencing then process the resultant amplified material to map out genomic variations in individual cells. Sequencing single sperm cells revealed higher mutation rates compared to the previous population measurements.⁹² This suggested genomic vulnerability within the first cellular divisions. Sequencing in single cancer cells revealed unique phases of tumour evolution based on copy number variation dynamics.⁹³ Bioinformatics analysis of CNVs within individual cells mapped out the phylogenetic trees of developing cells within organisms.⁹⁴ Somatic CNVs were detected in human brain health and disease.⁹⁵ Accumulation of somatic mutations in developing human neurons was also tracked using single cell DNA sequencing.⁹⁶

Single cell transcriptomics

Sequencing and imaging technologies have facilitated the quantification of mRNA molecules within a cellular volume. Unlike the fixed nature of DNA, RNA levels fluctuate in cells in response to stimuli. Thus, transcript measurements provide functionality of a single cell. Reverse transcription and whole-transcriptome amplification schemes enabled microarrays, quantitative PCR, and next generation sequencing (NGS) analysis for sequencing approaches at the single cell sensitivity. Transcripts from picograms (pg) of cellular RNA material are typically converted to complementary DNA (cDNA) during amplification to enable RNA sequencing read out.^{97,98} Recently, the reverse transcription step has been modified by an amplification step of a cDNA library together with adapter processing (tagmentation).⁹⁹ Despite this common cDNA usage feature, current single cell transcriptomics approaches differ based on their cell capturing strategies, targeting sites of a transcript (end vs. full length), quantitative counting capabilities (absolute or not), and strand specificity.¹⁰⁰ RNA sequencing methods have shed light on transcriptional codes within individual cells.

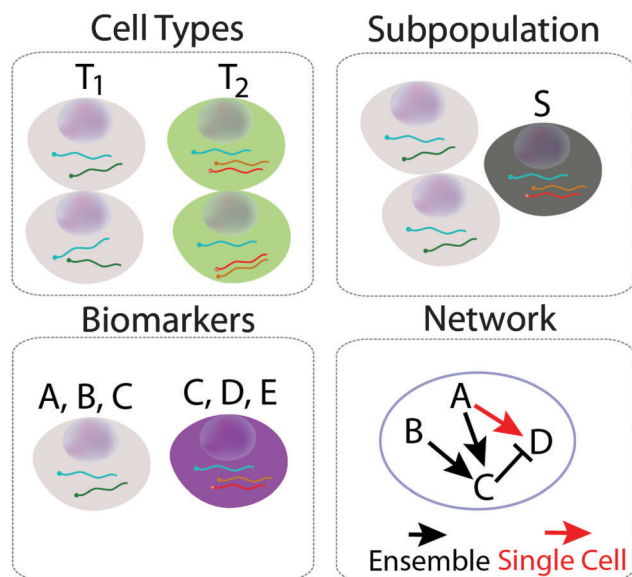


Fig. 4 Single cell profiling yields insights into cellular populations. These advantages include cell type discovery (T_1 and T_2 distinct cells with unique molecular content), subpopulation architecture identification (S denotes an emerging cellular subgroup), biomarker discovery ($A-C$ and $C-E$ are new identifiers of cells), and gene regulatory network refinement (red arrow shows previously undetected regulation partners).

These high-resolution transcriptional profiles of single cells reveal key population parameters such as the distinct cell types and subtypes, additional biomarkers per cell identity, and gene regulation modes observed in a new subpopulation (Fig. 4).

Cell type classification. While traditionally cells have recently been classified based on structure and function, the optimal cell type determination approach is to utilize underlying gene regulators and their connections.^{101,102} Therefore, single cell molecular profiles are paramount for the identification of distinct cell types within heterogeneous cell populations. Similarity in the transcriptional states of individual cells is utilized to extract common molecular signatures for a cell type. Generally, cells are classified by computational analyses of single cell RNA sequencing results using data presentation and reduction algorithms. Specifically, distinct clusters are obtained on hierarchical clustering heat map and principle component analysis plots, among many others. Currently, many cell-typing studies have been performed in one organ at a time.^{103–113} Using 96 to 44 000 and more cells, RNA sequencing methods have identified 3–47 and more cell types in different organs (Table 1). Covering all the organs would extend mapping of all the cell types within an organism to pave the way for a human atlas at cellular resolution. Besides, single-cell transcriptome studies provide differences in healthy and diseased cell states within individuals.

Biomarker discovery. Highly sensitive and specific biomarkers are needed to define the cellular states in health and disease.^{114,115} Together with conventional protein and nucleic acid labelling methods, DNA/RNA sequencing information about individual cells expands the toolset of testable molecular biomarkers. Single cell transcriptional profiles mapped out previously undetected biomarkers in the colon, the lungs,

Table 1 Transcriptional profiles of single cells are used to define cell identities and states in distinct organs. RNA sequencing identified similar cells based on computational analyses of transcriptional data. Previously unknown cell types, biomarkers, and developmental cues were observed from the single cell data

Specimen	Method	Throughput	Result
Colon ¹⁰³	PCR	96	5 cell types, 2 markers
Lungs ¹⁰⁴	RNA-seq	198	Cell marker identification
Brain ¹⁰⁵	RNA-seq	3000	47 subclasses
Intestine ¹⁰⁶	RNA-seq	238	3 subtypes, 1 marker
Spleen ¹⁰⁷	RNA-seq	4000	Immune cell types
Bone marrow ¹⁰⁸	RNA-seq	2730	Early commitment
Glioblastoma ¹⁰⁹	RNA-seq	430	Multiple subtypes
Tonsil ¹¹⁰	RNA-seq	648	Biomarker discovery
Neuron ¹¹¹	RNA-seq	622	11 cell types
Retina ¹¹²	Drop-seq	44 000	39 cell types
Stem cells ¹¹³	inDrop-seq	10 000	Gene network

the intestine, and the tonsil organs.^{103,104,106,110} In particular, the identification of biomarkers for immune cells significantly benefits from single-cell analysis. To identify markers, high expression levels of specific genes yielded distinct clusters in gene expression maps. For instance, transcriptional profiles of 4 distinct cell populations in the innate lymphoid cells (ILCs) revealed new markers in each cellular group such as GATA3 in ILC2 and RORC in ILC3 cells.¹¹⁰ These molecular identifiers are used for studies of homeostatic and inflammation conditions. Single cell transcriptional profiles allowed the identification of novel markers that were previously masked in population level measurements. Besides, single cell analysis of immune cells provides a T cell receptor repertoire that is developed against a specific infection.¹¹⁶ To map out the antigen diversity, the single cell transcriptomics approach yields clonally expanded T cell receptors for therapeutic applications and vaccination development.

Rewiring gene regulatory networks. Combinations of transcription factors interact with a genomic portion of DNA sequence clusters, and, at the same time, each transcription factor interacts with multiple genomic regions to regulate developmental processes, creating a gene regulatory network (GRN).¹¹⁷ Although genome-wide DNA/RNA sequencing provided significant GRN details, single cell transcriptional results have provided previously unrecognized regulatory relationships in the development of different cell types including blood cells.^{118–120} Single cell analysis has significant potential to dissect subpopulation architecture expressing combinations of genes that were obscured in the ensemble sequencing data. GRNs are much more complex when the cells differentiate into other cell types. Single cell analysis of 120 cells at 8 distinct time points during differentiation from human myeloid monocytic leukemia cells to macrophages revealed dynamic and specific rewiring.¹²¹

Noncoding transcript dynamics. Single cell transcriptional profiling by RNA sequencing revealed the dynamic role of lncRNAs in cellular programming. Specifically, lncRNA molecules regulate metabolic gene expression and repress lineage-specific genes.¹²² Single cell RNA fluorescent *in situ* hybridization (FISH) measurements on 61 distinct lncRNAs within three different cell types yielded abundance and subcellular localization patterns.¹²³

Most of the lncRNAs are localized to the nucleus with significant cell-to-cell variability, while some are spread across the cytoplasm similar to mRNA localization patterns. Another set of RNA sequencing measurements within individual oocytes showed that the lack of miRNA and small interfering RNA (siRNA) controlling genes (*Dicer1* and *Ago2*) abnormally upregulated another 1696 and 1553 genes, respectively.¹²⁴ Recently, single-cell analysis in the neocortex of the human brain revealed that lncRNAs are richly expressed in certain cell types in brain samples, which were previously masked by bulk studies.¹²⁵ These findings showed lncRNA's unique functions in specific cell types within the brain, creating distinct brain cell identities.

Transcript variant detection. RNA sequencing and FISH validation experiments on bone-marrow-derived individual dendritic cells provided heterogenous protein isoforms.¹²⁶ During the embryonic developmental stages, spanning the zygote to late blastocyst stages, unique splicing patterns were obtained per cell.¹²⁷ Two or more transcript isoforms of the same gene were observed in single cells, creating dynamic patterns of alternative splicing in human embryos and embryonic stem cells. Another statistical analysis method on RNA sequencing results showed the link between the cell cycle and alternative splicing.¹²⁸

Probing ribosome code. Dense transcripts for ribosomal proteins (typically found in the ribosomal subunits) were mapped out in cultures and thymus tissue sections, revealing cell type specific gene expression in single cells.¹²⁹ This observation supports the specialized ribosome theory as a cell identity regulator through the control of translational machinery. Two open questions would illuminate the ribosome code hypothesis at the single cell level. First, simultaneous profiling of ribosomal protein genes and metabolism related genes to test the role of the ribosome structure in transcript-specific translation within single cells. Next, mapping out directly ribosomal proteins (instead of the corresponding RNA molecules) to address the potential questions due to the short lifetime of RNAs. The RNA *versus* protein measurement concern holds for the single cell biology field in general, which should be taken into account also for cell identity studies.

Spatial mapping. While conventional RNA-seq achieves decent spatial resolution, imaging based *in situ* RNA profiling techniques provide spatial organization of cellular identities at the single-cell and single-molecule resolution. Spatially resolved RNA-seq (Tomo-seq) identified the role of BMP signaling in cardiomyocyte regeneration with coarse 3D mapping of a zebrafish heart, which was limited to large sample to sample collection distances.¹³⁰ On the other hand, *in situ* RNA sequencing approaches yield spatial distributions of molecules within a single cell or across multiple cells.^{131–134} Most *in situ* RNA detection methods are limited to flat cellular layers; however, signal amplification methods have been developed to screen transcriptional states of cellular types in thick tissue samples.^{135–137} The spatial arrangement of cells provides opportunities to study interactions of different cell types for the deconstruction of complex tissue and organ formation in development.

Single cell epigenomics

While most cellular identity results are from genomics and transcriptomics data, recent efforts in single cell epigenomics have started to shed light on cell type specific regulations. Single cell bisulfate sequencing has been developed for mapping DNA methylation across the genome within individual cells. Recent demonstrations revealed epigenetic diversity and dynamics in embryonic stem cells.^{138,139} Complementary imaging approaches screen a few epigenomic marks at a time. For instance, histone modifications have been detected by a proximity ligation assay in single cells. In particular, smooth muscle cells exhibited lineage specific dimethylation of lysine 4 of histone H3 (H3K4me2) in the *Myh11* locus in individual cells within tissue sections.¹⁴⁰ Live imaging of reporter assays provided *in vivo* regulation of DNA methylation patterns in individual cells.¹⁴¹ Another reporter assay showed the effect of epigenetic modifications on transcriptional activity.¹⁴² Another form of DNA methylation variant 5hmC (5-hydroxymethylcytosine) was recently sequenced at the single cell level. Based on the glucosylation-dependent digestion of DNA, 5hmC cell-to-cell variations were obtained and used for lineage reconstruction in mouse embryos.¹⁴³

Integration of advanced imaging, microfluidics, and barcoding enables a single cell ChIP-Seq method to map out chromatin states in individual cells. Despite its low coverage, different embryonic stem cells exhibited heterogeneous chromatin profiles.¹⁴⁴ Another low input ChIP-Seq was devised to measure histone mark profiles in individual cells.¹⁴⁵

Simultaneous measurements of the methylome and transcriptome of the same individual cells show the link between DNA methylation and transcription. Previously unrecognized heterogeneity of methylation in distal regulatory elements contributed to the transcriptional activity of pluripotency genes in embryonic stem cells (ESCs).¹⁴⁶ After obtaining substantial results compared to standalone bisulfite sequencing, this approach validated the correlation of the methylation patterns of non-CpG island (CGI) promoters with transcriptional repression. These results suggested significant epigenetic heterogeneity from cell to cell, especially in pluripotency factors such as *Esrrb* within serum ESCs.

Single cell proteomics

The protein content of an individual cell indicates its ultimate functional state. Powerful microscopy, flow cytometry,¹⁴⁷ and mass spectroscopy,¹⁴⁸ microarrays,¹⁴⁹ and immuno-PCR¹⁵⁰ techniques benefited from antibody labelled single cell assays to profile proteins. Antibody–oligonucleotide conjugate arrays within microfluidic chips enabled the highly multiplexed detection of proteins in single cells.^{151,152} Next generation sequencing enabled the detection of ligation assay based protein detection.^{153,154} Antibody tags exhibit specificity issues and are challenging to engineer for all proteins. To improve specific detection, the single cell western blot technique was developed to multiplex proteins within mammalian cells.¹⁵⁵ Rapid progress in proteomic profiling directs toward single cell proteomics.¹⁵⁶

Single cell chromosome-conformation-capture

As the physical conformation of the chromosome influences the gene expression and eventually the cell fate, the single-cell Hi-C technique is developed to provide a probe of the cell-to-cell organizational variation of the genome.^{157,158} Recently, the 4D Nucleome project, a National Institute of Health (NIH) funded consortium, has aimed to decipher the relationships between chromosomal conformation, tissue types, and diseases.

Physical mechanisms

The systems biology view of cell identity formation expanded our knowledge on coding mechanisms even down to single cell levels. However, cells make decisions under realistic physical factors to reach unique identities. To dissect the physical principles behind cellular specialization, we discuss the internal and external influencers that play a role in a cell's coordination and cooperation in development.

Intracellular

Molecular conversions within an individual cell are guided by physical rules. Investigation of the governing physical mechanisms in different cell types is crucial to understand how cellular identity is formed and properly maintained in health.

Kinetic proofreading. The genetic code encompasses specific recognition of transfer RNAs (tRNAs) to match anticodons to codons in mRNAs. The error rate in this process is 1 in 10 000, enabling correct protein synthesis practically all the time to form appropriate cellular identities. Despite the existence of other similar sized molecules, specific molecular interactions achieve proper assembly with low erroneous molecular recognition. Therefore, cells do not get affected by thermodynamic noise in gene translation owing to a kinetic proofreading (KP) mechanism. Proposed by John Hopfield, KP incorporates multiple selection steps to enrich correct binding. Incorrect conjugation still occurs but falls off before affecting the process.^{159,160} KP is also a general mechanism in cells to increase the specificity of molecular binding events including cell-to-cell interactions. For instance, blood cells utilize KP to recognize antigen-presenting cells in our immune system.

DNA–protein interactions. Different transcriptional programs in cells are regulated by interactions of proteins with DNA, leading to activation or repression of gene expression. Combinations of multiple transcription factors create cell specific transcriptional programming in both prokaryotes and eukaryotes.¹⁶¹ The binding events of TFs are often modelled by thermodynamics principles. A simple TF–DNA interaction models the total binding energies and binding probability of a TF to its target.¹⁶² The promoter architecture of each gene includes activators and repressors to regulate gene expression. A general formalism was developed to take into account the binding strength and abundance of these regulators. This approach linked physical parameters of transcriptional regulation fugacity to the fold-change (FC) in gene expression.¹⁶³

Chromatin structure. Another physical mechanism is that epigenetic codes change the gene expression by modifying the

structural distribution of chromatin.¹⁶⁴ Sequential histone modifications silence gene expression by compacting chromatin to form a heterochromatin.¹⁶⁵ A change in the structure of chromatin also allows the recruitment of other proteins.^{166,167} An analytical model was developed to investigate a few histone modifications.¹⁶⁸ DNA methylation increases the rigidity of the chromatin structure and disfavours the positioning of nucleosomes on DNA.^{169,170}

Splicing entropy. RNA maturation process includes the assembly of gene products through splicing machinery. Aberrant RNA types are formed in diseases such as cancer. This disorder is modelled by Shannon's entropy.^{171–173} Cancer cells exhibited highly significant splicing disorder compared to healthy subjects.

Extracellular

Cells make decisions under environmental factors such as signalling molecules, extracellular matrix (ECM) contexts, mechanical features, and interaction with other cells.¹⁷⁴ Cell's morphological characteristics such as shape, ECM's stiffness and topography are the main physical regulators to affect cell's identity development.¹⁷⁵ Stem cells sense the properties of an ECM by feedback signalling from mechanical and biochemical cues.¹⁷⁶ Engineering niches for stem cell studies will open up new directions in cellular reprogramming and tissue regeneration.¹⁷⁷

Applications

Research

Single cell approaches yield rich data to decipher complex biological networks in organisms with molecular sensitivity. These detailed cellular data are highly valuable for systems biology research. In particular, measuring gene products allows the development of cell specific mathematical models to describe biological information processing in single cells.^{178,179} These outcomes reveal a network of molecular components (nucleic acids and proteins) to formulate and test hypotheses in cellular machineries. Complementary implementation of single cell studies in medical settings facilitates systems medicine research to reconstruct biological networks in health and disease for translational purposes, paving the way for personalized and predictive medicine.^{180–183} These emerging directions provide opportunities to understand molecular principles of diseases and determine the best drug screening strategies in single cells. Primarily, cancer research leverages emerging single cell sequencing methods to study the molecular mechanisms of tumor heterogeneity, evolution, evasion, metastasis, and resistance.^{184,185} Single cell DNA/RNA sequencing in primary sites and circulating tumor cells revealed copy number variations, mutation rates, clonal dynamics, and transcriptional mapping within various cancer types including pancreatic, breast, lung, colon, and bladder cancers, and melanoma.

Diagnostics

Early screening and detection of diseases including cancer is critical to improve the healthcare system. Molecular profiling

techniques such as transcript profiling allowed biomarker discovery in disease to develop diagnostic assays. Bodily fluids such as human saliva were used as a non-invasive diagnostic method to define a set of biomarker signatures for early detection.¹⁸⁶ Transcriptomic biomarkers made it possible to differentiate lung cancer patients from normal subjects with high sensitivity and specificity. Similarly, a liquid biopsy method utilized single cell sequencing to profile circulating tumour cells in breast cancers.¹⁸⁷ Single cell genotyping revealed heterogenous mutations in acute myeloid leukaemia (AML).¹⁸⁸ In AML samples, deciphering the distribution of FLT3 and NPM1 mutations in clonal populations suggested significant tumor variability for better diagnosis and potentially for monitoring the progression of disease.

Therapeutics

The ultimate therapy should be to use patient derived induced pluripotent stem cells (iPS) for reprogramming to any cell type of interest towards a specific disorder. For instance, molecular changes during reprogramming from fibroblast to neurons have been dissected by single-cell RNA-seq.¹⁸⁹ This approach has been tested in medical settings for hematological, neurological, cardiovascular, metabolic, endocrine, and muscular disorders, yielding insights into drug based therapies.^{190,191} Reprogramming therapies included sickle cell anaemia treatment in mice by correcting mutations in hematopoietic stem cells followed by transplantation.¹⁹² Transferring iPS technology to correct human diseases is promising, but many challenges in complex molecular engineering efforts covering transcriptional and epigenetic regulations and gene editing approaches need to be addressed to realize stem cell based therapies.¹⁹³

Discussion

Multiple layers of regulators from chromatin states to post-translational modifications contribute to the cellular identity. The interconnection between these layers presents challenges to dissect the roles of each player. Hence, the ultimate goal is to merge studies of molecular decoding from different levels of gene control mechanisms. Simultaneous profiling of the genome, transcriptome, epigenome, and proteome has significant potential to reveal cell identity formation. Currently, the genome and transcriptome have been probed even in the same individual cell.^{194,195} Increasing efforts measure single cell correlations of transcripts and proteins.^{196,197}

Coverage in single cell sequencing is limited compared to bulk DNA/RNA sequencing due to amplification errors (allelic dropouts, distortion, false positive rates, and non-uniform usage) and small sample volumes. Mapping the entire genome (100% coverage) needs technological advancements to reduce sequence dropout errors, which can then accurately measure mutations without being affected by experimental noise. Increasing efforts such as MALBAC and MDA have improved the coverage levels of DNA sequencing to more than 90% in individual cells.^{198–202} On the other hand, current RNA sequencing methods have 5–25%

detection efficiency due to the poor conversion of RNAs to cDNA and eventually to amplified sequences. Reducing the reaction volume (from microliters to nanoliters within microfluidic chambers) captures more RNA molecules.⁹⁰ Alternative biochemical strategies are needed to enhance the RNA/cDNA conversion rate for efficient RNA sequencing approaches. Besides, to cover the entire proteome at the single cell level, efficient labelling strategies are desired to identify each and every protein in complex tissues and organs. Profiling other molecules (metabolites, lipids, small molecules, non-coding RNAs, among others) would complement existing single cell analysis techniques.

Conclusion

The presented cellular identity study will transform developmental biology research with a particular emphasis on the integration of systems level molecular analyses and underlying physical exploration. Understanding different levels of molecular coding at the single cell level will explain how cellular identity is gained in health and lost in disease. Physical control mechanisms will explain how cells specialize under many environmental stimulants. Finally, defining the cell state and reprogramming the cell identity will revolutionize medical research and practice.

Acknowledgements

A. F. C. is supported by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund. The authors thank Sten Linnarsson for help with figures.

Notes and references

- 1 E. Bianconi, A. Piovesan, F. Facchin, A. Beraudi, R. Casadei, F. Frabetti, L. Vitale, M. C. Pelleri, S. Tassani, F. Piva, S. Perez-Amodio, P. Strippoli and S. Canaider, *Ann. Hum. Biol.*, 2013, **40**, 463–471.
- 2 F. Crick, *Nature*, 1970, **227**, 561–563.
- 3 D. Greenbaum, C. Colangelo, K. Williams and M. Gerstein, *Genome Biol.*, 2003, **4**, 117.
- 4 E. S. Yeung, *Angew. Chem., Int. Ed.*, 2011, **50**, 583–585.
- 5 A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer and B. Wold, *Nat. Methods*, 2008, **5**, 621–628.
- 6 P. J. Park, *Nat. Rev. Genet.*, 2009, **10**, 669–680.
- 7 G. M. Rubin, M. D. Yandell, J. R. Wortman, G. L. Gabor Miklos, C. R. Nelson, I. K. Hariharan, M. E. Fortini, P. W. Li, R. Apweiler, W. Fleischmann, J. M. Cherry, S. Henikoff, M. P. Skupski, S. Misra, M. Ashburner, E. Birney, M. S. Boguski, T. Brody, P. Brokstein, S. E. Celniker, S. A. Chervitz, D. Coates, A. Cravchik, A. Gabrielian, R. F. Galle, W. M. Gelbart, R. A. George, L. S. B. Goldstein, F. Gong, P. Guan, N. L. Harris, B. A. Hay, R. A. Hoskins, J. Li, Z. Li, R. O. Hynes, S. J. M. Jones, P. M. Kuehl, B. Lemaitre, J. T. Littleton, D. K. Morrison, C. Mungall, P. H. O'Farrell, O. K. Pickeral, C. Shue,

- L. B. Vosshall, J. Zhang, Q. Zhao, X. H. Zheng, F. Zhong, W. Zhong, R. Gibbs, J. C. Venter, M. D. Adams and S. Lewis, *Science*, 2000, **287**, 2204–2215.
- 8 A. R. Mushegian, J. R. Garey, J. Martin and L. X. Liu, *Genome Res.*, 1998, **8**, 590–598.
 - 9 S. Bergmann, J. Ihmels and N. Barkai, *PLoS Biol.*, 2003, **2**, e9.
 - 10 A. T. Chinwalla, L. L. Cook, K. D. Delehaunty, G. A. Fewell, L. A. Fulton, R. S. Fulton, T. A. Graves, L. W. Hillier, E. R. Mardis, J. D. McPherson, T. L. Miner, W. E. Nash, J. O. Nelson, M. N. Nhan, K. H. Pepin, C. S. Pohl, T. C. Ponce, B. Schultz, J. Thompson, E. Trevaskis, R. H. Waterston, M. C. Wendl, R. K. Wilson, S.-P. Yang, P. An, E. Berry, B. Birren, T. Bloom, D. G. Brown, J. Butler, M. Daly, R. David, J. Deri, S. Dodge, K. Foley, D. Gage, S. Gnerre, T. Holzer, D. B. Jaffe, M. Kamal, E. K. Karlsson, C. Kells, A. Kirby, E. J. Kulbokas, E. S. Lander, T. Landers, J. P. Leger, R. Levine, K. Lindblad-Toh, E. Mauceli, J. H. Mayer, M. McCarthy, J. Meldrim, J. P. Mesirov, R. Nicol, C. Nusbaum, S. Seaman, T. Sharpe, A. Sheridan, J. B. Singer, R. Santos, B. Spencer, N. Stange-Thomann, J. P. Vinson, C. M. Wade, J. Wierzbowski, D. Wyman, M. C. Zody, E. Birney, N. Goldman, A. Kasprzyk, E. Mongin, A. G. Rust, G. Slater, A. Stabenau, A. Ureta-Vidal, S. Whelan, R. Ainscough, J. Attwood, J. Bailey, K. Barlow, S. Beck, J. Burton, M. Clamp, C. Clee, A. Coulson, J. Cuff, V. Curwen, T. Cutts, J. Davies, E. Eyra, D. Grafham, S. Gregory, T. Hubbard, A. Hunt, M. Jones, A. Joy, S. Leonard, C. Lloyd, L. Matthews, S. McLaren, K. McLay, B. Meredith, J. C. Mullikin, Z. Ning, K. Oliver, E. Overton-Larty, R. Plumb, S. Potter, M. Quail, J. Rogers, C. Scott, S. Searle, R. Shownkeen, S. Sims, M. Wall, A. P. West, D. Willey, S. Williams, J. F. Abril, R. Guigó, G. Parra, P. Agarwal, R. Agarwala, D. M. Church, W. Hlavina, D. R. Maglott, V. Sapozhnikov, M. Alexandersson, L. Pachter, S. E. Antonarakis, E. T. Dermitzakis, A. Reymond, C. Ucla, R. Baertsch, M. Diekhans, T. S. Furey, A. Hinrichs, F. Hsu, D. Karolchik, W. J. Kent, K. M. Roskin, M. S. Schwartz, C. Sugnet, R. J. Weber, P. Bork, I. Letunic, M. Suyama, D. Torrents, E. M. Zdobnov, M. Botcherby, S. D. Brown, R. D. Campbell, I. Jackson, N. Bray, O. Couronne, I. Dubchak, A. Poliakov, E. M. Rubin, M. R. Brent, P. Flicek, E. Keibler, I. Korf, S. Batalov, C. Bult, W. N. Frankel, P. Carninci, Y. Hayashizaki, J. Kawai, Y. Okazaki, S. Cawley, D. Kulp, R. Wheeler, F. Chiaromonte, F. S. Collins, A. Felsenfeld, M. Guyer, J. Peterson, K. Wetterstrand, R. R. Copley, R. Mott, C. Dewey, N. J. Dickens, R. D. Emes, L. Goodstadt, C. P. Ponting, E. Winter, D. M. Dunn, A. C. von Niederhausern, R. B. Weiss, S. R. Eddy, L. S. Johnson, T. A. Jones, L. Elnitski, D. L. Kolbe, P. Eswara, W. Miller, M. J. O'Connor, S. Schwartz, R. A. Gibbs, D. M. Muzny, G. Glusman, A. Smit, E. D. Green, R. C. Hardison, S. Yang, D. Haussler, A. Hua, B. A. Roe, R. S. Kucherlapati, K. T. Montgomery, J. Li, M. Li, S. Lucas, B. Ma, W. R. McCombie, M. Morgan, P. Pevzner, G. Tesler, J. Schultz, D. R. Smith, J. Tromp, K. C. Worley and E. D. Green, *Nature*, 2002, **420**, 520–562.
 - 11 J. F. Crow, *Daedalus*, 2002, **131**, 81–88.
 - 12 A. Kondrashov, *Nature*, 2012, **488**, 467–468.
 - 13 C. F. Baer, M. M. Miyamoto and D. R. Denver, *Nat. Rev. Genet.*, 2007, **8**, 619–631.
 - 14 A. Kong, M. L. Frigge, G. Masson, S. Besenbacher, P. Sulem, G. Magnusson, S. A. Gudjonsson, A. Sigurdsson, A. Jonasdottir, A. Jonasdottir, W. S. W. Wong, G. Sigurdsson, G. B. Walters, S. Steinberg, H. Helgason, G. Thorleifsson, D. F. Gudbjartsson, A. Helgason, O. T. Magnusson, U. Thorsteinsdottir and K. Stefansson, *Nature*, 2012, **488**, 471–475.
 - 15 F. H. C. Crick, *J. Mol. Biol.*, 1968, **38**, 367–379.
 - 16 A. Ambrogelly, S. Palioura and D. Söll, *Nat. Chem. Biol.*, 2007, **3**, 29–35.
 - 17 E. V. Koonin and A. S. Novozhilov, *IUBMB Life*, 2009, **61**, 99–111.
 - 18 D. Prandi, S. C. Baca, A. Romanel, C. E. Barbieri, J.-M. Mosquera, J. Fontugne, H. Beltran, A. Sboner, L. A. Garraway, M. A. Rubin and F. Demichelis, *Genome Biol.*, 2014, **15**, 439.
 - 19 J. Wang and R. E. Davis, *Curr. Opin. Genet. Dev.*, 2014, **27**, 26–34.
 - 20 T. S. Alioto, I. Buchhalter, S. Derdak, B. Hutter, M. D. Eldridge, E. Hovig, L. E. Heisler, T. A. Beck, J. T. Simpson, L. Tonon, A.-S. Sertier, A.-M. Patch, N. Jäger, P. Ginsbach, R. Drews, N. Paramasivam, R. Kabbe, S. Chotewutmontri, N. Diessl, C. Previti, S. Schmidt, B. Brors, L. Feuerbach, M. Heinold, S. Gröbner, A. Korshunov, P. S. Tarpey, A. P. Butler, J. Hinton, D. Jones, A. Menzies, K. Raine, R. Shepherd, L. Stebbings, J. W. Teague, P. Ribeca, F. C. Giner, S. Beltran, E. Raineri, M. Dabad, S. C. Heath, M. Gut, R. E. Denroche, N. J. Harding, T. N. Yamaguchi, A. Fujimoto, H. Nakagawa, V. Quesada, R. Valdés-Mas, S. Nakken, D. Vodák, L. Bower, A. G. Lynch, C. L. Anderson, N. Waddell, J. V. Pearson, S. M. Grimmond, M. Peto, P. Spellman, M. He, C. Kandath, S. Lee, J. Zhang, L. Létourneau, S. Ma, S. Seth, D. Torrents, L. Xi, D. A. Wheeler, C. López-Otín, E. Campo, P. J. Campbell, P. C. Boutros, X. S. Puente, D. S. Gerhard, S. M. Pfister, J. D. McPherson, T. J. Hudson, M. Schlesner, P. Lichter, R. Eils, D. T. W. Jones and I. G. Gut, *Nat. Commun.*, 2015, **6**, 10001.
 - 21 M. J. Keogh and P. F. Chinnery, *Clin. Neurol. Neurosurg.*, 2013, **115**, 948–953.
 - 22 E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander and J. Dekker, *Science*, 2009, **326**, 289–293.
 - 23 D. G. Lupiáñez, K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H. Kayserili, J. M. Opitz, R. Laxova, F. Santos-Simarro, B. Gilbert-Dussardier, L. Wittler,

- M. Borschiwer, S. A. Haas, M. Osterwalder, M. Franke, B. Timmermann, J. Hecht, M. Spielmann, A. Visel and S. Mundlos, *Cell*, 2015, **161**, 1012–1025.
- 24 D. G. Lupiáñez, M. Spielmann and S. Mundlos, *Trends Genet.*, 2016, **32**, 225–237.
- 25 P. H. L. Krijger, B. Di Stefano, E. de Wit, F. Limone, C. van Oevelen, W. de Laat and T. Graf, *Cell Stem Cell*, 2016, **18**, 597–610.
- 26 T. Marquardt and S. L. Pfaff, *Cell*, 2001, **106**, 651–654.
- 27 J. A. Zhang, A. Mortazavi, B. A. Williams, B. J. Wold and E. V. Rothenberg, *Cell*, 2012, **149**, 467–482.
- 28 S. Siegert, E. Cabuy, B. G. Scherf, H. Kohler, S. Panda, Y.-Z. Le, H. J. Fehling, D. Gaidatzis, M. B. Stadler and B. Roska, *Nat. Neurosci.*, 2012, **15**, 487–495.
- 29 C. Cobaleda, A. Schebesta, A. Delogu and M. Busslinger, *Nat. Immunol.*, 2007, **8**, 463–470.
- 30 T. I. Lee and R. A. Young, *Cell*, 2013, **152**, 1237–1251.
- 31 H. Sasaki-Iwaoka, K. Maruyama, H. Endoh, T. Komori, S. Kato and H. Kawashima, *J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res.*, 1999, **14**, 248–255.
- 32 Y. Zheng, C. Devitt, J. Liu, J. Mei and S. X. Skapek, *Dev. Biol.*, 2013, **380**, 49–57.
- 33 W. A. Whyte, D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee and R. A. Young, *Cell*, 2013, **153**, 307–319.
- 34 D. Hnisz, B. J. Abraham, T. I. Lee, A. Lau, V. Saint-André, A. A. Sigova, H. A. Hoke and R. A. Young, *Cell*, 2013, **155**, 934–947.
- 35 J. M. Downen, Z. P. Fan, D. Hnisz, G. Ren, B. J. Abraham, L. N. Zhang, A. S. Weintraub, J. Schuijers, T. I. Lee, K. Zhao and R. A. Young, *Cell*, 2014, **159**, 374–387.
- 36 D. Hay, J. R. Hughes, C. Babbs, J. O. J. Davies, B. J. Graham, L. L. P. Hanssen, M. T. Kassouf, A. M. Oudelaar, J. A. Sharpe, M. C. Suci, J. Telenius, R. Williams, C. Rode, P.-S. Li, L. A. Pennacchio, J. A. Sloane-Stanley, H. Ayyub, S. Butler, T. Sauka-Spengler, R. J. Gibbons, A. J. H. Smith, W. G. Wood and D. R. Higgs, *Nat. Genet.*, 2016, DOI: 10.1038/ng.3605.
- 37 B. M. Turner, *Nat. Cell Biol.*, 2007, **9**, 2–6.
- 38 M. Spivakov and A. G. Fisher, *Nat. Rev. Genet.*, 2007, **8**, 263–271.
- 39 B. D. Strahl and C. D. Allis, *Nature*, 2000, **403**, 41–45.
- 40 T. Jenuwein and C. D. Allis, *Science*, 2001, **293**, 1074–1080.
- 41 B. M. Turner, *Cell*, 2002, **111**, 285–291.
- 42 M. S. Cosgrove, J. D. Boeke and C. Wolberger, *Nat. Struct. Mol. Biol.*, 2004, **11**, 1037–1043.
- 43 N. D. Heintzman, G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenkov, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis and B. Ren, *Nature*, 2009, **459**, 108–112.
- 44 B. A. Benayoun, E. A. Pollina, D. Ucar, S. Mahmoudi, K. Karra, E. D. Wong, K. Devarajan, A. C. Daugherty, A. B. Kundaje, E. Mancini, B. C. Hitz, R. Gupta, T. A. Rando, J. C. Baker, M. P. Snyder, J. M. Cherry and A. Brunet, *Cell*, 2014, **158**, 673–688.
- 45 V. Orlando, *Cell*, 2003, **112**, 599–606.
- 46 T. I. Lee, R. G. Jenner, L. A. Boyer, M. G. Guenther, S. S. Levine, R. M. Kumar, B. Chevalier, S. E. Johnstone, M. F. Cole, K. Isono, H. Koseki, T. Fuchikami, K. Abe, H. L. Murray, J. P. Zucker, B. Yuan, G. W. Bell, E. Herbolzheimer, N. M. Hannett, K. Sun, D. T. Odom, A. P. Otte, T. L. Volkert, D. P. Bartel, D. A. Melton, D. K. Gifford, R. Jaenisch and R. A. Young, *Cell*, 2006, **125**, 301–313.
- 47 L. Ringrose and R. Paro, *Development*, 2007, **134**, 223–232.
- 48 A. Sparmann and M. van Lohuizen, *Nat. Rev. Cancer*, 2006, **6**, 846–856.
- 49 A. Bird, *Genes Dev.*, 2002, **16**, 6–21.
- 50 K. E. Varley, J. Gertz, K. M. Bowling, S. L. Parker, T. E. Reddy, F. Pauli-Behn, M. K. Cross, B. A. Williams, J. A. Stamatoyannopoulos, G. E. Crawford, D. M. Absher, B. J. Wold and R. M. Myers, *Genome Res.*, 2013, **23**, 555–567.
- 51 S. Dhawan, S. Georgia, S. Tschen, G. Fan and A. Bhushan, *Dev. Cell*, 2011, **20**, 419–429.
- 52 B. Bodega and V. Orlando, *Curr. Opin. Cell Biol.*, 2014, **31**, 67–73.
- 53 M. Bundo, M. Toyoshima, Y. Okada, W. Akamatsu, J. Ueda, T. Nemoto-Miyauchi, F. Sunaga, M. Toritsuka, D. Ikawa, A. Kakita, M. Kato, K. Kasai, T. Kishimoto, H. Nawa, H. Okano, T. Yoshikawa, T. Kato and K. Iwamoto, *Neuron*, 2014, **81**, 306–313.
- 54 G. D. Evrony, X. Cai, E. Lee, L. B. Hills, P. C. Elhosary, H. S. Lehmann, J. J. Parker, K. D. Atabay, E. C. Gilmore, A. Poduri, P. J. Park and C. A. Walsh, *Cell*, 2012, **151**, 483–496.
- 55 G. J. Faulkner, Y. Kimura, C. O. Daub, S. Wani, C. Plessy, K. M. Irvine, K. Schroder, N. Cloonan, A. L. Steptoe, T. Lassmann, K. Waki, N. Hornig, T. Arakawa, H. Takahashi, J. Kawai, A. R. R. Forrest, H. Suzuki, Y. Hayashizaki, D. A. Hume, V. Orlando, S. M. Grimmond and P. Carninci, *Nat. Genet.*, 2009, **41**, 563–571.
- 56 A. R. Muotri, M. C. N. Marchetto, N. G. Coufal, R. Oefner, G. Yeo, K. Nakashima and F. H. Gage, *Nature*, 2010, **468**, 443–446.
- 57 S. Wissing, M. Muñoz-Lopez, A. Macia, Z. Yang, M. Montano, W. Collins, J. L. Garcia-Perez, J. V. Moran and W. C. Greene, *Hum. Mol. Genet.*, 2012, **21**, 208–218.
- 58 S. Loewer, M. N. Cabili, M. Guttman, Y.-H. Loh, K. Thomas, I. H. Park, M. Garber, M. Curran, T. Onder, S. Agarwal, P. D. Manos, S. Datta, E. S. Lander, T. M. Schlaeger, G. Q. Daley and J. L. Rinn, *Nat. Genet.*, 2010, **42**, 1113–1117.
- 59 M. Guttman, I. Amit, M. Garber, C. French, M. F. Lin, D. Feldser, M. Huarte, O. Zuk, B. W. Carey, J. P. Cassady, M. N. Cabili, R. Jaenisch, T. S. Mikkelsen, T. Jacks, N. Hacohen, B. E. Bernstein, M. Kellis, A. Regev, J. L. Rinn and E. S. Lander, *Nature*, 2009, **458**, 223–227.
- 60 A. Fatica and I. Bozzoni, *Nat. Rev. Genet.*, 2014, **15**, 7–21.

- 61 S. Quinodoz and M. Guttman, *Trends Cell Biol.*, 2014, **24**, 651–663.
- 62 T. Xia, Q. Liao, X. Jiang, Y. Shao, B. Xiao, Y. Xi and J. Guo, *Sci. Rep.*, 2014, **4**, 6088.
- 63 P. Paci, T. Colombo and L. Farina, *BMC Syst. Biol.*, 2014, **8**, 83.
- 64 F. Gaiti, S. L. Fernandez-Valverde, N. Nakanishi, A. D. Calcino, I. Yanai, M. Tanurdzic and B. M. Degnan, *Mol. Biol. Evol.*, 2015, msv117.
- 65 Z. Du, T. Sun, E. Hacisuleyman, T. Fei, X. Wang, M. Brown, J. L. Rinn, M. G.-S. Lee, Y. Chen, P. W. Kantoff and X. S. Liu, *Nat. Commun.*, 2016, **7**, 10982.
- 66 A. O. Ribeiro, C. R. G. Schoof, A. Izzotti, L. V. Pereira and L. R. Vasques, *MicroRNA*, 2014, **3**, 45–53.
- 67 H. Kaspi, R. Pasvolsky and E. Hornstein, *Trends Endocrinol. Metab.*, 2014, **25**, 285–292.
- 68 M. C. Siomi, K. Sato, D. Pezic and A. A. Aravin, *Nat. Rev. Mol. Cell Biol.*, 2011, **12**, 246–258.
- 69 J. Gonzalez, H. Qi, N. Liu and H. Lin, *Cell Rep.*, 2015, **12**, 150–161.
- 70 X. Ma, S. Wang, T. Do, X. Song, M. Inaba, Y. Nishimoto, L. Liu, Y. Gao, Y. Mao, H. Li, W. McDowell, J. Park, K. Malanowski, A. Peak, A. Perera, H. Li, K. Gaudenz, J. Haug, Y. Yamashita, H. Lin, J. Ni and T. Xie, *PLoS One*, 2014, **9**, e90267.
- 71 N. Charlet-B, G. Singh, T. A. Cooper and P. Logan, *Mol. Cell*, 2002, **9**, 649–658.
- 72 C. J. David and J. L. Manley, *Genes Dev.*, 2008, **22**, 279–285.
- 73 A. J. Matlin, F. Clark and C. W. J. Smith, *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 386–398.
- 74 M. Gabut, P. Samavarchi-Tehrani, X. Wang, V. Slobodeniuc, D. O'Hanlon, H.-K. Sung, M. Alvarez, S. Talukder, Q. Pan, E. O. Mazzone, S. Nedelec, H. Wichterle, K. Woltjen, T. R. Hughes, P. W. Zandstra, A. Nagy, J. L. Wrana and B. J. Blencowe, *Cell*, 2011, **147**, 132–146.
- 75 X. Yang, J. Coulombe-Huntington, S. Kang, G. M. Sheynkman, T. Hao, A. Richardson, S. Sun, F. Yang, Y. A. Shen, R. R. Murray, K. Spirohn, B. E. Begg, M. Duran-Frigola, A. MacWilliams, S. J. Pevzner, Q. Zhong, S. A. Trigg, S. Tam, L. Ghamsari, N. Sahni, S. Yi, M. D. Rodriguez, D. Balcha, G. Tan, M. Costanzo, B. Andrews, C. Boone, X. J. Zhou, K. Salehi-Ashtiani, B. Charleaux, A. A. Chen, M. A. Calderwood, P. Aloy, F. P. Roth, D. E. Hill, L. M. Iakoucheva, Y. Xia and M. Vidal, *Cell*, 2016, **164**, 805–817.
- 76 J. S. Mattick and M. F. Mehler, *Trends Neurosci.*, 2008, **31**, 227–233.
- 77 J. J. C. Rosenthal and P. H. Seeburg, *Neuron*, 2012, **74**, 432–439.
- 78 E. Park, B. Williams, B. J. Wold and A. Mortazavi, *Genome Res.*, 2012, **22**, 1626–1633.
- 79 Q. Zhang and X. Xiao, *Nat. Methods*, 2015, **12**, 347–350.
- 80 S. Sharma, S. K. Patnaik, R. Thomas Taggart, E. D. Kannisto, S. M. Enriquez, P. Gollnick and B. E. Baysal, *Nat. Commun.*, 2015, **6**, 6881.
- 81 S. Xue, S. Tian, K. Fujii, W. Kladwang, R. Das and M. Barna, *Nature*, 2015, **517**, 33–38.
- 82 S. Xue and M. Barna, *Nat. Rev. Mol. Cell Biol.*, 2012, **13**, 355–369.
- 83 N. Slavov, S. Semrau, E. Airoidi, B. Budnik and A. van Oudenaarden, *Cell Rep.*, 2015, **13**, 865–873.
- 84 M. Gallego and D. M. Virshup, *Nat. Rev. Mol. Cell Biol.*, 2007, **8**, 139–148.
- 85 S. Westermann and K. Weber, *Nat. Rev. Mol. Cell Biol.*, 2003, **4**, 938–948.
- 86 A. M. Bode and Z. Dong, *Nat. Rev. Cancer*, 2004, **4**, 793–805.
- 87 C. M. Dobson, *Nature*, 2003, **426**, 884–890.
- 88 J. P. Thaler, S.-K. Lee, L. W. Jurata, G. N. Gill and S. L. Pfaff, *Cell*, 2002, **110**, 237–249.
- 89 Q. F. Wills, K. J. Livak, A. J. Tipping, T. Enver, A. J. Goldson, D. W. Sexton and C. Holmes, *Nat. Biotechnol.*, 2013, **31**, 748–752.
- 90 S. De, *Trends Genet. TIG*, 2011, **27**, 217–223.
- 91 M. J. McConnell, M. R. Lindberg, K. J. Brennand, J. C. Piper, T. Voet, C. Cowing-Zitron, S. Shumilina, R. S. Lasken, J. R. Vermeesch, I. M. Hall and F. H. Gage, *Science*, 2013, **342**, 632–637.
- 92 J. Wang, H. C. Fan, B. Behr and S. R. Quake, *Cell*, 2012, **150**, 402–412.
- 93 N. Navin, J. Kendall, J. Troge, P. Andrews, L. Rodgers, J. McIndoo, K. Cook, A. Stepansky, D. Levy, D. Esposito, L. Muthuswamy, A. Krasnitz, W. R. McCombie, J. Hicks and M. Wigler, *Nature*, 2011, **472**, 90–94.
- 94 T. Garvin, R. Aboukhalil, J. Kendall, T. Baslan, G. S. Atwal, J. Hicks, M. Wigler and M. C. Schatz, *Nat. Methods*, 2015, **12**, 1058–1060.
- 95 X. Cai, G. D. Evrony, H. S. Lehmann, P. C. Elhosary, B. K. Mehta, A. Poduri and C. A. Walsh, *Cell Rep.*, 2014, **8**, 1280–1289.
- 96 M. A. Lodato, M. B. Woodworth, S. Lee, G. D. Evrony, B. K. Mehta, A. Karger, S. Lee, T. W. Chittenden, A. M. D'Gama, X. Cai, L. J. Luquette, E. Lee, P. J. Park and C. A. Walsh, *Science*, 2015, **350**, 94–98.
- 97 S. Islam, A. Zeisel, S. Joost, G. La Manno, P. Zajac, M. Kasper, P. Lönnerberg and S. Linnarsson, *Nat. Methods*, 2014, **11**, 163–166.
- 98 S. Islam, U. Kjällquist, A. Moliner, P. Zajac, J.-B. Fan, P. Lönnerberg and S. Linnarsson, *Nat. Protoc.*, 2012, **7**, 813–828.
- 99 S. Picelli, Å. K. Björklund, B. Reinius, S. Sagasser, G. Winberg and R. Sandberg, *Genome Res.*, 2014, **24**, 2033–2040.
- 100 L. Wen and F. Tang, *Genome Biol.*, 2016, **17**, 71.
- 101 D. Arendt, *Nat. Rev. Genet.*, 2008, **9**, 868–882.
- 102 M. K. Vickaryous and B. K. Hall, *Biol. Rev. Cambridge Philos. Soc.*, 2006, **81**, 425–455.
- 103 P. Dalerba, T. Kalisky, D. Sahoo, P. S. Rajendran, M. E. Rothenberg, A. A. Leyrat, S. Sim, J. Okamoto, D. M. Johnston, D. Qian, M. Zabala, J. Bueno, N. F. Neff, J. Wang, A. A. Shelton, B. Visser, S. Hisamori, Y. Shimono, M. van de Wetering, H. Clevers, M. F. Clarke and S. R. Quake, *Nat. Biotechnol.*, 2011, **29**, 1120–1127.
- 104 B. Treutlein, D. G. Brownfield, A. R. Wu, N. F. Neff, G. L. Mantalas, F. H. Espinoza, T. J. Desai, M. A. Krasnow and S. R. Quake, *Nature*, 2014, **509**, 371–375.

- 105 A. Zeisel, A. B. Muñoz-Manchado, S. Codeluppi, P. Lönnerberg, G. L. Manno, A. Juréus, S. Marques, H. Munguba, L. He, C. Betsholtz, C. Rolny, G. Castelo-Branco, J. Hjerling-Leffler and S. Linnarsson, *Science*, 2015, **347**, 1138–1142.
- 106 D. Grün, A. Lyubimova, L. Kester, K. Wiebrands, O. Basak, N. Sasaki, H. Clevers and A. van Oudenaarden, *Nature*, 2015, **525**, 251–255.
- 107 D. A. Jaitin, E. Kenigsberg, H. Keren-Shaul, N. Elefant, F. Paul, I. Zaretsky, A. Mildner, N. Cohen, S. Jung, A. Tanay and I. Amit, *Science*, 2014, **343**, 776–779.
- 108 F. Paul, Y. Arkin, A. Giladi, D. A. Jaitin, E. Kenigsberg, H. Keren-Shaul, D. Winter, D. Lara-Astiaso, M. Gury, A. Weiner, E. David, N. Cohen, F. K. B. Lauridsen, S. Haas, A. Schlitzer, A. Mildner, F. Ginhoux, S. Jung, A. Trumpp, B. T. Porse, A. Tanay and I. Amit, *Cell*, 2015, **163**, 1663–1677.
- 109 A. P. Patel, I. Tirosh, J. J. Trombetta, A. K. Shalek, S. M. Gillespie, H. Wakimoto, D. P. Cahill, B. V. Nahed, W. T. Curry, R. L. Martuza, D. N. Louis, O. Rozenblatt-Rosen, M. L. Suvà, A. Regev and B. E. Bernstein, *Science*, 2014, 1254257.
- 110 Å. K. Björklund, M. Forkel, S. Picelli, V. Konya, J. Theorell, D. Friberg, R. Sandberg and J. Mjösberg, *Nat. Immunol.*, 2016, **17**, 451–460.
- 111 D. Usoskin, A. Furlan, S. Islam, H. Abdo, P. Lönnerberg, D. Lou, J. Hjerling-Leffler, J. Haeggström, O. Kharchenko, P. V. Kharchenko, S. Linnarsson and P. Ernfors, *Nat. Neurosci.*, 2015, **18**, 145–153.
- 112 E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev and S. A. McCarroll, *Cell*, 2015, **161**, 1202–1214.
- 113 A. M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz and M. W. Kirschner, *Cell*, 2015, **161**, 1187–1201.
- 114 N. Rifai, M. A. Gillette and S. A. Carr, *Nat. Biotechnol.*, 2006, **24**, 971–983.
- 115 H. Runne, A. Kuhn, E. J. Wild, W. Pratyaksha, M. Kristiansen, J. D. Isaacs, E. Régulier, M. Delorenzi, S. J. Tabrizi and R. Luthi-Carter, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 14424–14429.
- 116 A. Han, J. Glanville, L. Hansmann and M. M. Davis, *Nat. Biotechnol.*, 2014, **32**, 684–692.
- 117 M. Levine and E. H. Davidson, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 4936–4942.
- 118 V. Moignard, I. C. Macaulay, G. Swiers, F. Buettner, J. Schütte, F. J. Calero-Nieto, S. Kinston, A. Joshi, R. Hannah, F. J. Theis, S. E. Jacobsen, M. F. de Bruijn and B. Göttgens, *Nat. Cell Biol.*, 2013, **15**, 363–372.
- 119 V. Moignard, S. Woodhouse, L. Haghverdi, A. J. Lilly, Y. Tanaka, A. C. Wilkinson, F. Buettner, I. C. Macaulay, W. Jawaid, E. Diamanti, S.-I. Nishikawa, N. Piterman, V. Kouskoff, F. J. Theis, J. Fisher and B. Göttgens, *Nat. Biotechnol.*, 2015, **33**, 269–276.
- 120 C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N. J. Lennon, K. J. Livak, T. S. Mikkelsen and J. L. Rinn, *Nat. Biotechnol.*, 2014, **32**, 381–386.
- 121 T. Kouno, M. de Hoon, J. C. Mar, Y. Tomaru, M. Kawano, P. Carninci, H. Suzuki, Y. Hayashizaki and J. W. Shin, *Genome Biol.*, 2013, **14**, R118.
- 122 D. H. Kim, G. K. Marinov, S. Pepke, Z. S. Singer, P. He, B. Williams, G. P. Schroth, M. B. Elowitz and B. J. Wold, *Cell Stem Cell*, 2015, **16**, 88–101.
- 123 M. N. Cabili, M. C. Dunagin, P. D. McClanahan, A. Biais, O. Padovan-Merhar, A. Regev, J. L. Rinn and A. Raj, *Genome Biol.*, 2015, **16**, 20.
- 124 F. Tang, C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B. B. Tuch, A. Siddiqui, K. Lao and M. A. Surani, *Nat. Methods*, 2009, **6**, 377–382.
- 125 Q. Ma and H. Y. Chang, *Genome Biol.*, 2016, **17**, 68.
- 126 A. K. Shalek, R. Satija, X. Adiconis, R. S. Gertner, J. T. Gaub, R. Raychowdhury, S. Schwartz, N. Yosef, C. Malboeuf, D. Lu, J. J. Trombetta, D. Gennert, A. Gnirke, A. Goren, N. Hacohen, J. Z. Levin, H. Park and A. Regev, *Nature*, 2013, **498**, 236–240.
- 127 L. Yan, M. Yang, H. Guo, L. Yang, J. Wu, R. Li, P. Liu, Y. Lian, X. Zheng, J. Yan, J. Huang, M. Li, X. Wu, L. Wen, K. Lao, R. Li, J. Qiao and F. Tang, *Nat. Struct. Mol. Biol.*, 2013, **20**, 1131–1139.
- 128 J. D. Welch, Y. Hu and J. F. Prins, *Nucleic Acids Res.*, 2016, gkv1525.
- 129 A. F. Coskun and L. Cai, *Nat. Methods*, 2016, DOI: 10.1038/nmeth.3895.
- 130 C.-C. Wu, F. Kruse, M. D. Vasudevarao, J. P. Junker, D. C. Zebrowski, K. Fischer, E. S. Noël, D. Grün, E. Berezikov, F. B. Engel, A. van Oudenaarden, G. Weidinger and J. Bakkers, *Dev. Cell*, 2016, **36**, 36–49.
- 131 E. Lubeck, A. F. Coskun, T. Zhiyentayev, M. Ahmad and L. Cai, *Nat. Methods*, 2014, **11**, 360–361.
- 132 J. H. Lee, E. R. Daugherty, J. Scheiman, R. Kalhor, J. L. Yang, T. C. Ferrante, R. Terry, S. S. F. Jeanty, C. Li, R. Amamoto, D. T. Peters, B. M. Turczyk, A. H. Marblestone, S. A. Inverso, A. Bernard, P. Mali, X. Rios, J. Aach and G. M. Church, *Science*, 2014, **343**, 1360–1363.
- 133 K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang and X. Zhuang, *Science*, 2015, **348**, aaa6090.
- 134 N. Crosetto, M. Bienko and A. van Oudenaarden, *Nat. Rev. Genet.*, 2015, **16**, 57–66.
- 135 H. M. T. Choi, J. Y. Chang, L. A. Trinh, J. E. Padilla, S. E. Fraser and N. A. Pierce, *Nat. Biotechnol.*, 2010, **28**, 1208–1212.
- 136 R. Ke, M. Mignardi, A. Pacureanu, J. Svedlund, J. Botling, C. Wählby and M. Nilsson, *Nat. Methods*, 2013, **10**, 857–860.
- 137 P. L. Ståhl, F. Salmén, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, A. Mollbrink, S. Linnarsson, S. Codeluppi, Å. Borg, F. Pontén, P. I. Costea, P. Sahlén, J. Mulder, O. Bergmann, J. Lundeberg and J. Frisén, *Science*, 2016, **353**, 78–82.

- 138 M. Farlik, N. C. Sheffield, A. Nuzzo, P. Datlinger, A. Schönegger, J. Klughammer and C. Bock, *Cell Rep.*, 2015, **10**, 1386–1397.
- 139 S. A. Smallwood, H. J. Lee, C. Angermueller, F. Krueger, H. Saadeh, J. Peat, S. R. Andrews, O. Stegle, W. Reik and G. Kelsey, *Nat. Methods*, 2014, **11**, 817–820.
- 140 D. Gomez, L. S. Shankman, A. T. Nguyen and G. K. Owens, *Nat. Methods*, 2013, **10**, 171–177.
- 141 Y. Stelzer, C. S. Shivalila, F. Soldner, S. Markoulaki and R. Jaenisch, *Cell*, 2015, **163**, 218–229.
- 142 L. Bintu, J. Yong, Y. E. Antebi, K. McCue, Y. Kazuki, N. Uno, M. Oshimura and M. B. Elowitz, *Science*, 2016, **351**, 720–724.
- 143 D. Mooijman, S. S. Dey, J.-C. Boisset, N. Crosetto and A. van Oudenaarden, *Nat. Biotechnol.*, 2016, DOI: 10.1038/nbt.3598.
- 144 A. Rotem, O. Ram, N. Shores, R. A. Sperling, A. Goren, D. A. Weitz and B. E. Bernstein, *Nat. Biotechnol.*, 2015, **33**, 1165–1172.
- 145 J. Brind'Amour, S. Liu, M. Hudson, C. Chen, M. M. Karimi and M. C. Lorincz, *Nat. Commun.*, 2015, **6**, 6033.
- 146 C. Angermueller, S. J. Clark, H. J. Lee, I. C. Macaulay, M. J. Teng, T. X. Hu, F. Krueger, S. A. Smallwood, C. P. Ponting, T. Voet, G. Kelsey, O. Stegle and W. Reik, *Nat. Methods*, 2016, **13**, 229–232.
- 147 S. P. Perfetto, P. K. Chattopadhyay and M. Roederer, *Nat. Rev. Immunol.*, 2004, **4**, 648–655.
- 148 S. C. Bendall, E. F. Simonds, P. Qiu, E. D. Amir, P. O. Krutzik, R. Finck, R. V. Bruggner, R. Melamed, A. Trejo, O. I. Ornatsky, R. S. Balderas, S. K. Plevritis, K. Sachs, D. Pe'er, S. D. Tanner and G. P. Nolan, *Science*, 2011, **332**, 687–696.
- 149 G. MacBeath, *Nat. Genet.*, 2002, **32**, 526–532.
- 150 C. M. Niemeyer, M. Adler and R. Wacker, *Nat. Protoc.*, 2007, **2**, 1918–1930.
- 151 R. Fan, O. Vermesh, A. Srivastava, B. K. H. Yen, L. Qin, H. Ahmad, G. A. Kwong, C.-C. Liu, J. Gould, L. Hood and J. R. Heath, *Nat. Biotechnol.*, 2008, **26**, 1373–1378.
- 152 Q. Shi, L. Qin, W. Wei, F. Geng, R. Fan, Y. S. Shin, D. Guo, L. Hood, P. S. Mischel and J. R. Heath, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 419–424.
- 153 S. Darmanis, R. Y. Nong, J. Vänelid, A. Siegbahn, O. Ericsson, S. Fredriksson, C. Bäcklin, M. Gut, S. Heath, I. G. Gut, L. Wallentin, M. G. Gustafsson, M. Kamali-Moghaddam and U. Landegren, *PLoS One*, 2011, **6**, e25583.
- 154 S. Fredriksson, W. Dixon, H. Ji, A. C. Koong, M. Mindrinos and R. W. Davis, *Nat. Methods*, 2007, **4**, 327–329.
- 155 A. J. Hughes, D. P. Spelke, Z. Xu, C.-C. Kang, D. V. Schaffer and A. E. Herr, *Nat. Methods*, 2014, **11**, 749–755.
- 156 A. F. M. Altelaar, J. Munoz and A. J. R. Heck, *Nat. Rev. Genet.*, 2013, **14**, 35–48.
- 157 T. Nagano, Y. Lubling, T. J. Stevens, S. Schoenfelder, E. Yaffe, W. Dean, E. D. Laue, A. Tanay and P. Fraser, *Nature*, 2013, **502**, 59–64.
- 158 O. Schwartzman and A. Tanay, *Nat. Rev. Genet.*, 2015, **16**, 716–726.
- 159 J. J. Hopfield, *Proc. Natl. Acad. Sci. U. S. A.*, 1974, **71**, 4135–4139.
- 160 M. Johansson, J. Zhang and M. Ehrenberg, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 131–136.
- 161 N. E. Buchler, U. Gerland and T. Hwa, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 5136–5141.
- 162 U. Gerland, J. D. Moroz and T. Hwa, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12015–12020.
- 163 F. M. Weinert, R. C. Brewster, M. Rydenfelt, R. Phillips and W. K. Kegel, *Phys. Rev. Lett.*, 2014, **113**, 258101.
- 164 A. J. Bannister and T. Kouzarides, *Cell Res.*, 2011, **21**, 381–395.
- 165 S. I. S. Grewal and D. Moazed, *Science*, 2003, **301**, 798–802.
- 166 T. Kouzarides, *Cell*, 2007, **128**, 693–705.
- 167 F. Andreoli and A. Del Rio, *Drug Discovery Today*, 2014, **19**, 1372–1379.
- 168 W. L. Ku, M. Girvan, G.-C. Yuan, F. Sorrentino and E. Ott, *PLoS One*, 2013, **8**, e77944.
- 169 T. I. Yusufaly, Y. Li and W. K. Olson, *J. Phys. Chem. B*, 2013, **117**, 16436–16442.
- 170 A. Pérez, C. L. Castellazzi, F. Battistini, K. Collinet, O. Flores, O. Deniz, M. L. Ruiz, D. Torrents, R. Eritja, M. Soler-López and M. Orozco, *Biophys. J.*, 2012, **102**, 2140–2148.
- 171 M. Osella and M. Caselle, *Phys. Biol.*, 2009, **6**, 046018.
- 172 W. Ritchie, S. Granjeaud, D. Puthier and D. Gautheret, *PLoS Comput. Biol.*, 2008, **4**, e1000011.
- 173 G. Yeo and C. B. Burge, *J. Comput. Biol.*, 2004, **11**, 377–394.
- 174 F. M. Watt and W. T. S. Huck, *Nat. Rev. Mol. Cell Biol.*, 2013, **14**, 467–473.
- 175 F. Guilak, D. M. Cohen, B. T. Estes, J. M. Gimble, W. Liedtke and C. S. Chen, *Cell Stem Cell*, 2009, **5**, 17–26.
- 176 H. Lv, L. Li, M. Sun, Y. Zhang, L. Chen, Y. Rong and Y. Li, *Stem Cell Res. Ther.*, 2015, **6**, 103.
- 177 S. W. Lane, D. A. Williams and F. M. Watt, *Nat. Biotechnol.*, 2014, **32**, 795–803.
- 178 T. Ideker, T. Galitski and L. Hood, *Annu. Rev. Genomics Hum. Genet.*, 2001, **2**, 343–372.
- 179 H. Kitano, *Science*, 2002, **295**, 1662–1664.
- 180 C. Auffray, Z. Chen and L. Hood, *Genome Med.*, 2009, **1**, 2.
- 181 L. Hood and Q. Tian, *Genomics, Proteomics Bioinf.*, 2012, **10**, 181–185.
- 182 A. Mardinoglu and J. Nielsen, *J. Intern. Med.*, 2012, **271**, 142–154.
- 183 J. Nielsen, *J. Intern. Med.*, 2012, **271**, 108–110.
- 184 N. E. Navin, *Genome Res.*, 2015, **25**, 1499–1507.
- 185 P. Van Loo and T. Voet, *Curr. Opin. Genet. Dev.*, 2014, **24**, 82–91.
- 186 L. Zhang, H. Xiao, H. Zhou, S. Santiago, J. M. Lee, E. B. Garon, J. Yang, O. Brinkmann, X. Yan, D. Akin, D. Chia, D. Elashoff, N.-H. Park and D. T. W. Wong, *Cell. Mol. Life Sci.*, 2012, **69**, 3341–3350.
- 187 A. A. Powell, A. H. Talasz, H. Zhang, M. A. Coram, A. Reddy, G. Deng, M. L. Telli, R. H. Advani, R. W. Carlson, J. A. Mollick, S. Sheth, A. W. Kurian, J. M. Ford, F. E. Stockdale, S. R. Quake, R. F. Pease, M. N. Mindrinos, G. Bhanot, S. H. Dairkee, R. W. Davis and S. S. Jeffrey, *PLoS One*, 2012, **7**, e33788.
- 188 A. L. Paguirigan, J. Smith, S. Meshinchi, M. Carroll, C. Maley and J. P. Radich, *Sci. Transl. Med.*, 2015, **7**, 281re2.

- 189 B. Treutlein, Q. Y. Lee, J. G. Camp, M. Mall, W. Koh, S. A. M. Shariati, S. Sim, N. F. Neff, J. M. Skotheim, M. Wernig and S. R. Quake, *Nature*, 2016, **534**, 391–395.
- 190 A. B. C. Cherry and G. Q. Daley, *Cell*, 2012, **148**, 1110–1122.
- 191 D. A. Robinton and G. Q. Daley, *Nature*, 2012, **481**, 295–305.
- 192 D. R. Higgs, *N. Engl. J. Med.*, 2008, **358**, 964–966.
- 193 S. A. Goldman, *Cell Stem Cell*, 2016, **18**, 174–188.
- 194 S. S. Dey, L. Kester, B. Spanjaard, M. Bienko and A. van Oudenaarden, *Nat. Biotechnol.*, 2015, **33**, 285–289.
- 195 I. C. Macaulay, W. Haerty, P. Kumar, Y. I. Li, T. X. Hu, M. J. Teng, M. Goolam, N. Saurat, P. Coupland, L. M. Shirley, M. Smith, N. Van der Aa, R. Banerjee, P. D. Ellis, M. A. Quail, H. P. Swerdlow, M. Zernicka-Goetz, F. J. Livesey, C. P. Ponting and T. Voet, *Nat. Methods*, 2015, **12**, 519–522.
- 196 S. Darmanis, C. J. Gallant, V. D. Marinescu, M. Niklasson, A. Segerman, G. Flamourakis, S. Fredriksson, E. Assarsson, M. Lundberg, S. Nelander, B. Westermark and U. Landegren, *Cell Rep.*, 2016, **14**, 380–389.
- 197 C. Albayrak, C. A. Jordi, C. Zechner, J. Lin, C. A. Bichsel, M. Khammash and S. Tay, *Mol. Cell*, 2016, **61**, 914–924.
- 198 T. Daley and A. D. Smith, *Bioinformatics*, 2014, **30**, 3159–3165.
- 199 A. A. Pollen, T. J. Nowakowski, J. Shuga, X. Wang, A. A. Leyrat, J. H. Lui, N. Li, L. Szpankowski, B. Fowler, P. Chen, N. Ramalingam, G. Sun, M. Thu, M. Norris, R. Lebofsky, D. Toppani, D. W. Kemp II, M. Wong, B. Clerkson, B. N. Jones, S. Wu, L. Knutsson, B. Alvarado, J. Wang, L. S. Weaver, A. P. May, R. C. Jones, M. A. Unger, A. R. Kriegstein and J. A. A. West, *Nat. Biotechnol.*, 2014, **32**, 1053–1058.
- 200 D. Sims, I. Sudbery, N. E. Ilott, A. Heger and C. P. Ponting, *Nat. Rev. Genet.*, 2014, **15**, 121–132.
- 201 D. Grün and A. van Oudenaarden, *Cell*, 2015, **163**, 799–810.
- 202 A. M. Streets and Y. Huang, *Nat. Biotechnol.*, 2014, **32**, 1005–1006.